

Multiclass classification $x_i = \text{datapoint}$ $y_i = \text{subset of labels}$
 L classes: $\{1, \dots, L\}$

	Labels					
	1	2	3		L	
x_1	1	1	0	0	0	y_1 multi-label
x_2	0	1	0	0	0	y_2
x_3	0	1	0	1	0	

Approaches

I independent ~ Train a classifier separately for each label (binary).
 - test independently.

II Establish a priori label dependency.

Example $l=2, 3$ indep; $l=4$ depends on 2; label 1 depends on 2, 4
 label 5 depends on 3, 4



train classifiers by dependency IN ORDER, same for prediction

$$\bullet f_2(x) \simeq y_2 \quad ; \quad f_3(x) \simeq y_3$$

"classifier chain"

$$\bullet f_4(x, y_2) \simeq y_4$$

$$\bullet f_1(x, y_2, y_4) \simeq y_1$$

$$\bullet f_5(x, y_3, y_4) \simeq y_5$$

• can work well if dependencies of labels are established in adv
(medicine diagnosis)
(topology classification)
(hierarchy)

III Powerset

metabel $\boxed{Y = \text{subset category}}$

$\Rightarrow$ potentially 2^L possible metabels.
example movie cat $\begin{cases} 1 = \text{romance} & 4 = \text{drama} \\ 2 = \text{drama} & 5 = \text{war} \\ 3 = \text{comedy} \end{cases}$
 $y = \text{metabel} = \begin{pmatrix} \text{doc} \\ \text{war} \end{pmatrix}$

use multiclass (one out of many) classifier for metabels as categories.

train (x, y_{cat})

OK if 2^L labels, still some metabels sparse

IV Error Correcting Output Codes.

encode classes $\Rightarrow$ codes (binary)

encode multiclass $\Rightarrow$ codes labels

$k \approx \frac{\# \text{ classes}}{\# \text{ combinations (subsets) observed in training}} \ll 2^L$

binary classifiers

class labels	b_1	b_2	code word b_k					
1	0	1	0	1	0	1	0	1
2	0	1	1	0	0	0	1	0
3	1	0	0	1	0	0	0	0
4								
5	1	1	1	1	0	0	0	0

all bits = binary

multiclass classification $\Rightarrow$ predict one class out of many.

$\text{pred}(z)$:

0111 0000

b_k separately.

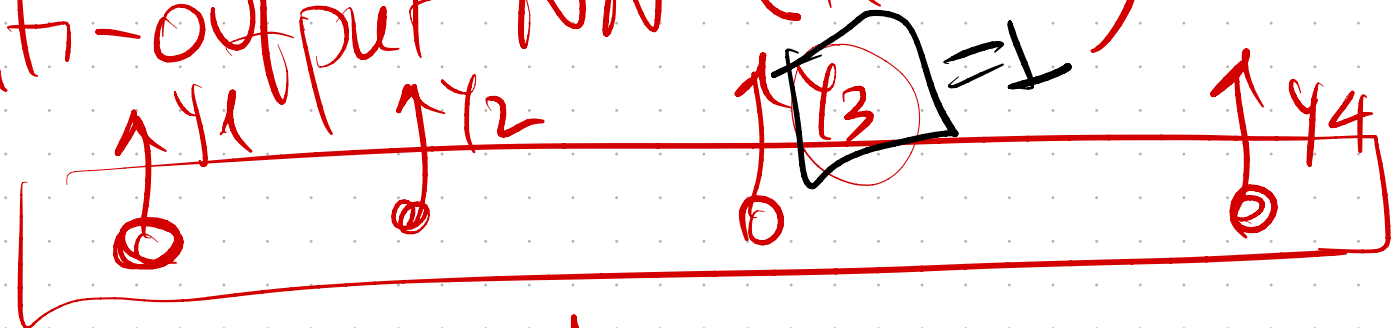
combinations (classes)

~~train binary classifiers (k) for $b_1, b_2, \dots, b_k$~~

test on datapoint z : $b_1(z) \ b_2(z) \ \dots \ b_k(z)$ *code word* predictions

look up the class with code word $\approx$ predicted code word.

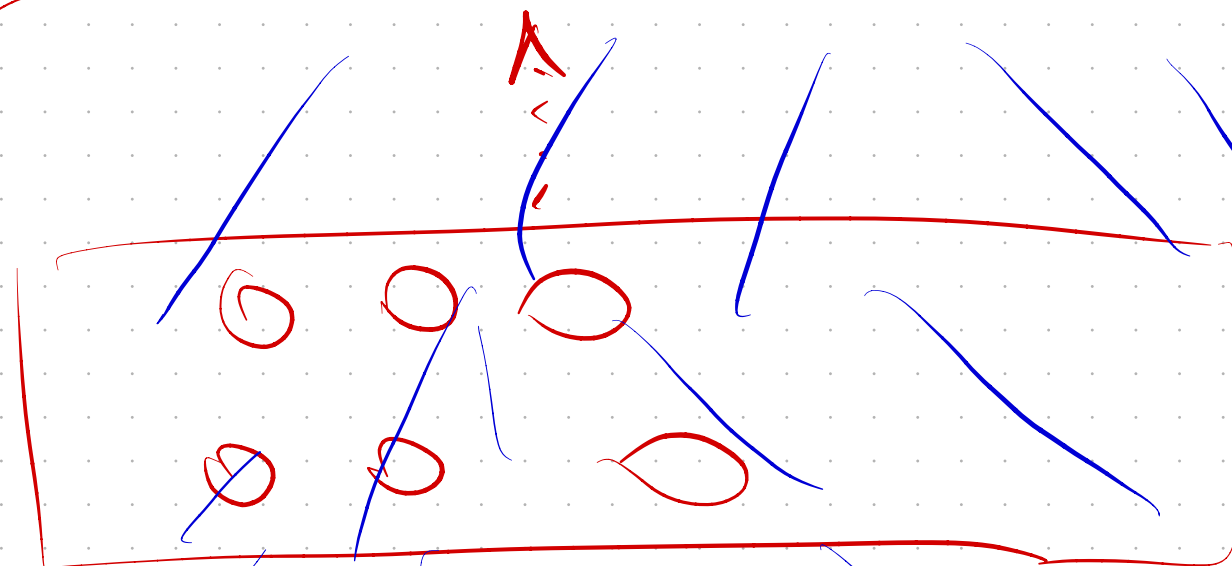
Multi-output NN (HW 2B)



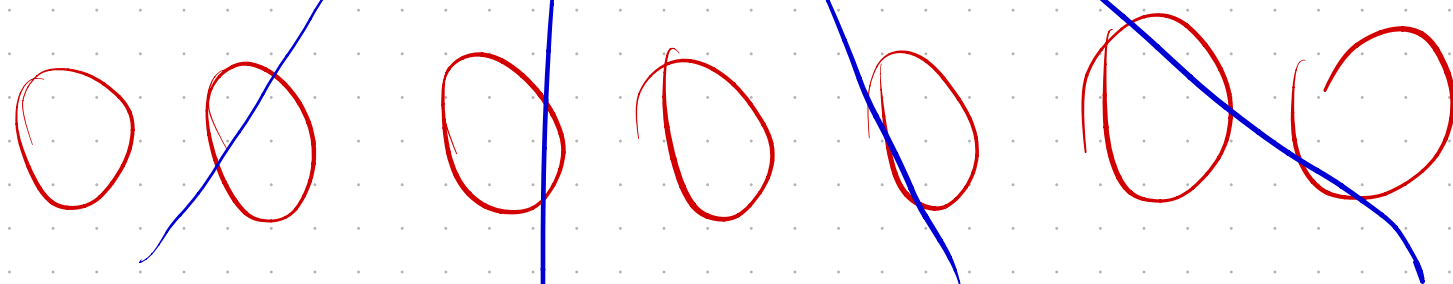
output

• multiclass
one of $y=1$, rest $=0$

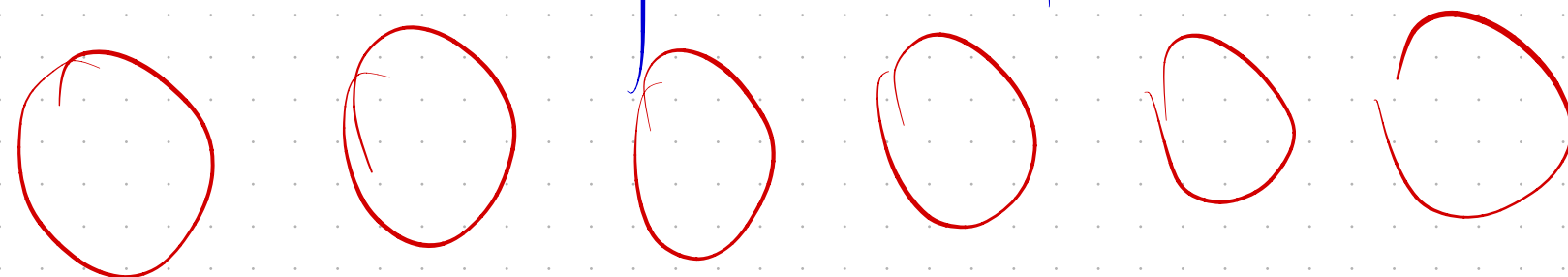
• multilabel:
 $y=1$ for all labels
in the subset



hidden



feat



input

sharing weights,
neurons output

label dependency